

# Analyzing Twitter Data to Evaluate the People's Attitudes to Public Health Policies and Events in the Era of COVID-19

Meng-Hsiu Tsai

University of Tennessee at Chattanooga

Yingfeng Wang (✉ [ywang802@gmail.com](mailto:ywang802@gmail.com))

University of Tennessee at Chattanooga <https://orcid.org/0000-0002-3715-5124>

---

## Research article

**Keywords:** Sentiment analysis, Social media, Recursive neural network

**Posted Date:** August 12th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-48957/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Analyzing Twitter Data to Evaluate the People's Attitudes to Public Health Policies and Events in the Era of COVID-19

Meng-Hsiu Tsai<sup>1</sup>, Yingfeng Wang<sup>1,\*</sup>

1. Department of Computer Science and Engineering, University of Tennessee at Chattanooga, Chattanooga, TN 37403, United States

\* Correspondence: yingfeng-wang@utc.edu

## Abstract

**Background:** In the era of a pandemic like COVID-19, monitoring the sentimental changes of the population is an urgent need, especially for the policy makers of the public health. A possible solution is to build a fast and low-cost surveillance system by using the sentiment analysis of Twitter data. Unfortunately, choosing a suitable sentiment classification model is still challenging. The general pre-trained model may be insensitive to the new specific terms of the pandemic. The early-trained model may have a bias issue due to the incomplete specific corpus. Although it is reasonable to assume the late-trained model is relatively reliable, it is usually available months after a pandemic begins.

**Methods:** This paper conducts the sentiment analysis of Twitter data and compares different models. Furthermore, we propose a strategy for using the pre-trained, early-trained, and late-trained models in a surveillance system based on Twitter data. The first two models can be used together in the early stage, while the last model can be used in the late stage. This study also analyzes the relationship between the sentimental changes of COVID-19-related Twitter data and the public health policies and events.

**Results:** Our results indicate that applying the pre-trained model to preprocessing early training samples may improve the early-trained model. Both models can work together by making up each other in the surveillance system in the early stage.

**Conclusions:** A fast and low-cost surveillance system is critical to the policy makers of the public health in a pandemic. This work uses the sentiment analysis of Twitter data to evaluate people's attitudes to public health policies and events. We propose a strategy to make the surveillance system effective since the early stage. This study also connects the sentimental changes of COVID-19-related Twitter data to the public health policies and events.

**Keywords:** Sentiment analysis, Social media, Recursive neural network

## 1. Introduction

The importance of policy-relevant evidence for evaluating the potential impact of public health policies has been widely recognized [1–3]. The practice of evidence-based public health policy (EBPH) aims to promote public health by fully utilizing available data for decision making [4]. Public health scientists have developed many approaches in this field [5–8]. The success of applying EBPH strongly encourages researchers to advance these methods [2, 9–11]. One common key component of these methods is evidence collection, which is still a challenging step [3]. The evidence data preparation is required to fit the policy design window [12, 13]. Also, policy surveillance systems need to monitor patterns and trends of the related policy influence [14, 15]. These systems require the data collection is not only efficient but also cheap. Unfortunately, traditional approaches cannot completely meet these requirements due to the inherent limitation of their evidence collection approaches [16]. In the era of a pandemic, e.g., COVID-19, the need for fast and low-cost evidence collection approaches is even more urgent.

To address the above challenges, scientists attempt to develop evidence collection approaches based on data from social media [17]. The ease of accessing the large amount of real-time social media data gives social media unique advantages for EBPH [16–20]. The dynamic social media platforms enable scientists to study large populations [21]. Furthermore, the real-time nature of social media allows fast data processing to avoid time delay caused by conventional approaches [16, 22]. These features are crucial to EBPH, especially in the era of a pandemic. Among these social media platforms, Twitter has been identified as a popular platform for short messages [18]. Each Twitter message, called tweet, is currently limited to contain up to 280 characters. The setting meets the need for quick and short updates and makes Twitter more and more popular. As of the first quarter of 2020, Twitter, a microblogging web service, has 166 million daily active users [23], who send more than 500 million tweets every day [24]. The broad demographic breadth and interactive nature of Twitter attracts researchers to apply it in EBPH [18, 25].

Among many tweet analysis areas, tweet sentiment analysis has already been a hot area. Sentiment analysis is a computational tool of natural language processing for studying the attitudes of the public on a topic [26, 27]. Recently, machine learning approaches have made encouraging progresses in sentiment analysis [28, 29]. Twitter provides a huge volume of tweets, most of which are unstructured public short text messages. This enables applying sentiment analysis of tweets in multiple areas [30]. Some of these applications addressed public health issues. These studies usually used a supervised learning machine, e.g., support vector machine (SVM), to analyze the sentiment of tweets. There are two ways to build the training data set. One way is the topic-independent strategy. Ji et al. measured the concerns of public health by counting the number of personal negative tweets of related topics every day [31]. They used the pre-specified positive and negative data to establish the training data set for their machine learning method. Another way is the topic-dependent strategy. Du et al. conducted the sentiment analysis on tweets related to human papillomavirus (HPV) vaccines [32] [33]. They manually annotated tweets and used these annotated tweets as the training data set. Cole-Lewis et al. conducted the sentiment analysis on tweets for studying people's attitudes to e-cigarettes [34]. In their machine learning method, they also made the training data set by manually annotating tweets. Daniulaityte et al. investigated drug-related tweets by sentiment analysis [35]. They manually labeled thousands of tweets for building the training data set of their machine learning method. Whether training the model by manually annotated tweets on specific topics is critical to a real-time policy surveillance system. The training based on manually annotated data may provide higher accuracy. However, the corpus of a specific public health topic is usually incomplete in the early stage. This disadvantage may lower the reliability of the early-trained models. If we frequently conduct manual annotation for a surveillance system, it may significantly increase the cost of this system and cause time delay. This issue is addressed in this paper.

The COVID-19 pandemic causes wide impacts and brings a significant number of related tweets. It allows us to investigate epidemic-related tweets and contribute to the public health. The major contribution of this work is summarized as follows,

- 1) This work investigates the tweets related to COVID-19 and compares sentiment analysis results with the pre-trained model and models trained by manually annotated tweets on COVID-19. Our investigation shows both approaches are useful and can work together.
- 2) Based on the above comparison, we propose a feasible strategy for building an almost real-

time surveillance system to evaluate the overall attitudes of the public by conducting sentiment analysis. It is reasonable to assume that the late-trained model, which is trained by sample tweets covering multiple months, is more reliable than the early-trained model, which is trained by early sample tweets. However, the late-trained model is unavailable in the early stage. This strategy uses the pre-trained model to improve the early-trained model by adjusting the training data set composition.

- 3) This work also analyzes the impacts of some COVID-19-related public policies and events by using the sentiment analysis of tweets.

The rest of this paper is organized as follows. Section 2 describes the methods of analyzing tweets and comparing different models. Section 3 proposes a new strategy for using multiple models together in a surveillance system. Section 4 analyzes sentimental scores of COVID-19-related tweets and connects the peaks and valleys of the scores to the public policies and events. The discussion is given in section 5. Section 6 concludes this paper.

## 2. Analysis and Comparison Methods

The brief pipeline of investigating related tweets and comparing different models is outlined in Figure 1. Details of this pipeline are given in the rest of this section.

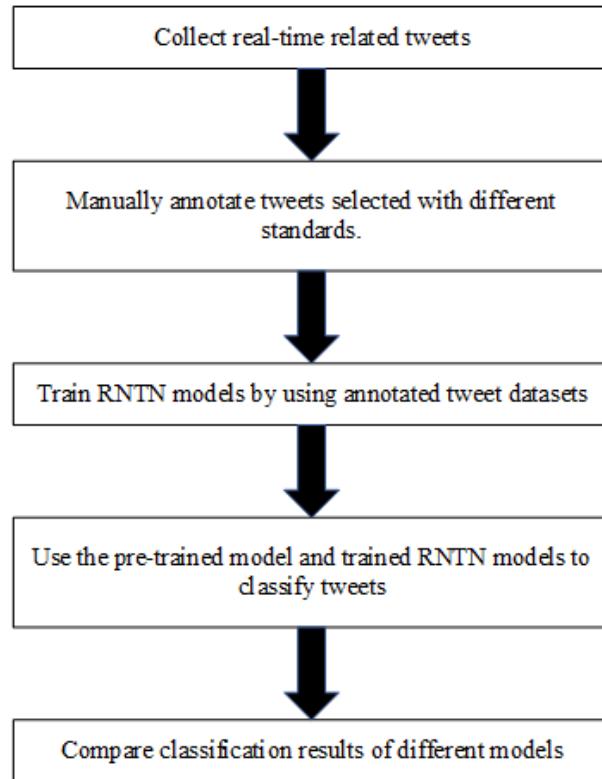


Fig.1. The brief pipeline of investigating related tweets and comparing different models.

### 2.1 Data Collection

We used Tweepy [36], a free python library for accessing twitter APIs, to collect real-time tweets. From March 1st, 2020 to June 14th, 2020, we collected real-time tweets for hours at noon every third day. Additionally, we collected related tweets at noon on March 12th due to the Europe travel ban announced by the United States government on March 11th. There are 37 data collection days. In each collection day, we received about 40,000 original tweets with searching key words "coronavirus", "COVID-19", "COVID19", "COVID\_19", "SARSCOV2", "SARS-COV-2", and "SARS\_COV\_2".

## 2.2 Data Preprocessing

We randomly select 700 tweets collected on March 1st. All 700 tweets and all words of a tweet are manually scored. Scores 0, 1, 2, 3, and 4 refer to very negative, negative, neutral, positive, and very positive, respectively. The annotated data set is called Mar01. Similarly, we randomly select 700 tweets collected in March and manually score them. The annotated data set is called March. Then, we repeat the same approach on tweets collected from March to May. The related dataset is called MarAprMay. Each data set is composed of 700 annotated tweets.

## 2.3 Model Training

In this paper, we use the Stanford CoreNLP toolkit [37] for sentiment classification. This is a pipeline framework for natural language processing (NLP). We use its recursive neural tensor network (RNTN) [38], a special type of recursive neural network (RNN), to conduct the sentiment analysis of tweets. The toolkit parses a preprocessed tweet into a binary tree. Each leaf node of this tree refers to a word. RNTN follows the bottom-up order to compute internal nodes by using a compositional function. Moreover, it classifies each node into one of the five sentimental categories that are very negative, negative, neutral, positive, and very positive. The sentiment of the root node is that of the whole tweet. The current model has already been trained and tuned but can also be trained by user-specified training data sets. We use data sets Mar01, March, and MarAprMay to train RNTN separately. All training parameters are set by default. The number of training samples in a batch is 27. The training can repeat up to 400 iterations. The learning rate is 0.01. As for each dataset, 600 tweets are assigned in the training data set, while 100 tweets are assigned to the validation data set.

## 2.4 Tweet classification

We use the pre-trained model and all three trained models to classify all collected tweets into scores 0, 1, 2, 3, and 4, which refer to very negative, negative, neutral, positive, and very positive, respectively. As for each model, the average score of a day is the overall sentimental score of that day. The way of computing the overall sentimental score for the  $i$ th collection day, represented by  $S_i$ , is given as follows.

$$S_i = \frac{\text{Sum(sentimental scores of all tweets on day } i)}{\text{Number of tweets collected on day } i}, \quad (1)$$

where  $\text{Sum}()$  is the summation function, while  $i$  is an integer between 1 and 37. The overall sentimental scores reflect the overall attitudes of the population. All overall sentimental scores

based on all four models are given in Figure 2. To further analyze these scores, we need to investigate the trends of sentimental score changes. It tells us whether the majority of the population are becoming more optimistic or panic. The sentimental score change for the  $i$ th collection day, represented by  $\Delta S_i$ , is given as follows.

$$\Delta S_i = \frac{S_i - S_1}{S_1} , \quad (2)$$

where  $i$  is an integer between 1 and 37. It is worth noting that some small shifts could be caused by noise. It requires us to filter out these shifts by smoothing the score changes. To meet this need, we may calculate the smoothed score change,  $\overline{\Delta S}_i$ , by the following expression.

$$\overline{\Delta S}_i = \begin{cases} \Delta S_i, & i = 1, 37 \\ (\Delta S_{i-1} + \Delta S_i + \Delta S_{i+1})/3, & 1 < i < 37 \end{cases} , \quad (3)$$

where  $i$  is an integer between 1 and 37. Please note that smoothing score changes may shift peaks and valleys to their neighboring days, because formula (3) also considers neighboring collection days. All smoothed score changes based on all models are given in Figure 3.

## 2.5 Model Comparison

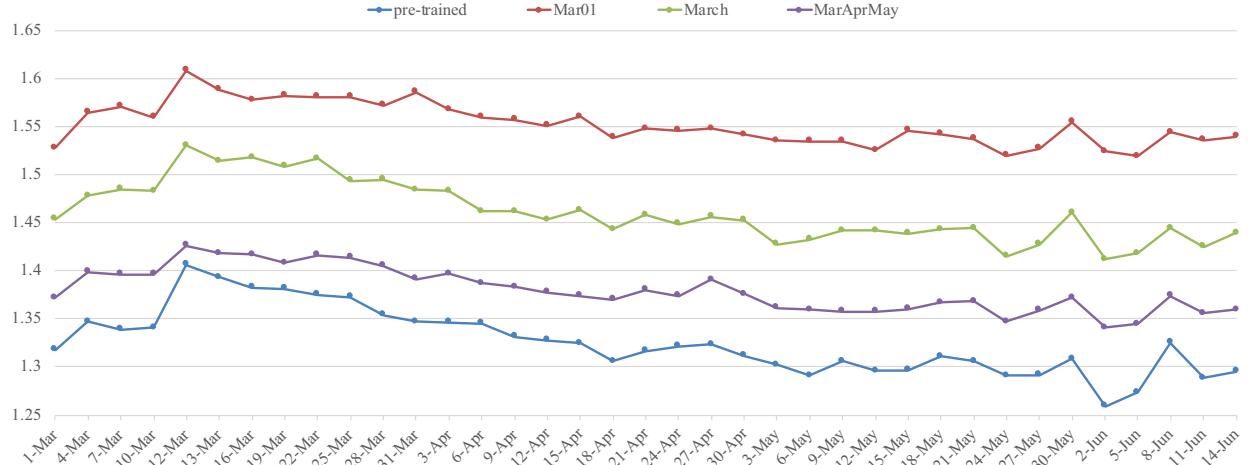


Fig. 2. The overall sentimental scores of the pre-trained model, model Mar01, model March, and model MarAprMay.

It is reasonable to assume that the model trained by sample tweets covering multiple months is more reliable than the model trained by early sample tweets. Therefore, we assume model MarAprMay is the most reliable trained model. Figure 2 shows overall sentimental scores based on all four models. It indicates the scores of models Mar01 and March are consistently greater than that of model MarAprMay, while scores of the pre-trained model are consistently less than that of model MarAprMay. Furthermore, scores of all models follow the same big trends. It indicates that the pre-trained model and trained models can be used to evaluate sentimental changes in a surveillance system roughly. Also, these two types of models can make up each other. Figure 3 shows smoothed score changes of all four models. It indicates that these models do not always agree with others. If we decide to use model MarAprMay, which is considered to be relatively

reliable, the surveillance system has to wait until June. This challenge motivates us to develop a strategy for improving the reliability of models trained by early sample tweets.

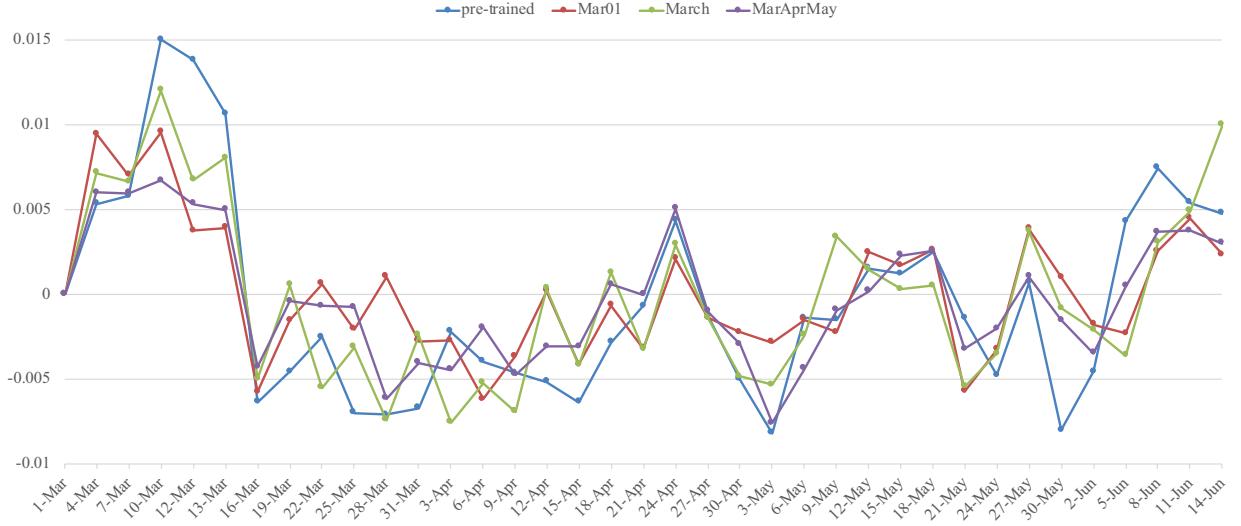


Fig. 3. The smoothed score changes of the pre-trained model, model Mar01, model March, and model MarAprMay.

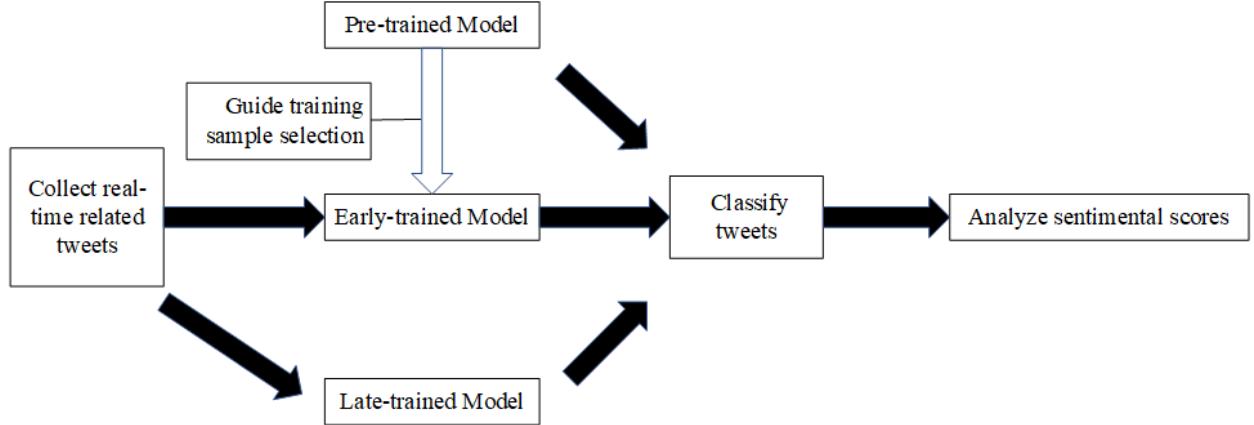


Fig. 4. The pipeline of the proposed strategy for the surveillance system. This system uses the pre-trained and early-trained models in the early stage and the late-trained model in the late stage.

### 3. Strategy for Multiple Models

All 700 training samples for model Mar01 were collected on March 1st. If its reliability is improved, we will use this model in the surveillance system. Given all annotated tweets are early samples, the data set may already have inherent biases due to the incompleteness of the corpus of COVID-19. Therefore, the tolerance of additional bias is relatively low. For example, the training set has 35, 331, 212, 115, and 7 tweets with sentimental scores 0, 1, 2, 3, and 4, respectively. The unbalanced sample composition may have worse impacts on Mar01 than other models.

Here, we use the pre-trained model to calibrate the training set of Mar01. We randomly select some tweets scored 3 or 4 by the pre-trained model. After manually checking these tweets are positive

or very positive, we add these tweets to the training set and remove the same number of negative or very negative tweets. We repeatedly update the training set until it is balanced. To verify the effectiveness of this method, we use the training data set for MarAprMay as the testing data set for Mar01 and the new model, named M01Update. The overall sentimental scores based on the manual annotation, Mar01, and M01Update are 1.68, 1.53, and 1.72, respectively. The overall sentimental score of M01Update is closer to that of manual annotation. We also calculate the mean squared error (MSE) by the following formula,

$$MSE = \frac{\sum_{k=1}^n (Score_k^{manual} - Score_k^{model})^2}{n}, \quad (4)$$

where  $n$  is the number of tweets in the data set, that is, 700 in our tests, while  $Score_k^{manual}$  and  $Score_k^{model}$  refer to the scores of  $k$ th tweet based on manual annotation and model classification, respectively. A smaller MSE indicates the better performance. The MSE for Mar01 is 1.20, while the MSE for M01Update is 1.10. It also confirms that M01Update performs better than Mar01.

The above progress enables us to use the pre-trained model to guide training sample selection for model M01Update, which is an early-trained model. Both models can work together by making up each other in the surveillance system in the early stage. When the late-trained model, e.g., the model trained by annotated tweets covering three months, is available, it can provide more reliable sentimental scores for the surveillance system, while the pre-trained and early-trained models can still generate complementary analysis results. Moreover, this surveillance system only requires manual annotation for two data sets. One is for the early-trained model, the other is for the late-trained model. The summarized pipeline is described in Figure 4. This system fully utilizes the pre-trained, early-trained, and late-trained models, so it can consistently conduct fast, low-cost, and relatively reliable sentiment analysis.

#### 4. Sentimental Score Analysis

In this section, we study the sentimental results produced by MarAprMay for analyzing the impacts of the COVID-19 related public policies and events. Figure 5 shows the results of collected tweets from March 1st to June 14th. According to the stacked columns in this figure, the percentages of very positive and very negative are consistently low, while the percentage changes of positive are also insignificant. The major reasons of score changes are the sentiment switches between neural and negative. It suggests that the majority of the population is panic about COVID-19.

Figure 5 shows some visible sharp score peaks. The peak on March 12th reflects the positive score change due to the Europe travel ban announced by the United States government on March 11th [39]. This sharp peak may also suggests that the Twitter users from the United States may be the major influence of the sentimental score changes. The peak on April 27th happened when many states of the United States announced or decided to end the stay-at-home order soon. This order started at the end of March or in the beginning of April and might be related to the continuous negative sentimental score changes [40]. The widely reported rapid development of COVID-19 vaccine may be the reason leading to the peak on May 30th [41]. Similarly, the widely reported positive effects achieved by the shutdown policies may also bring the peak on June 8th [42].

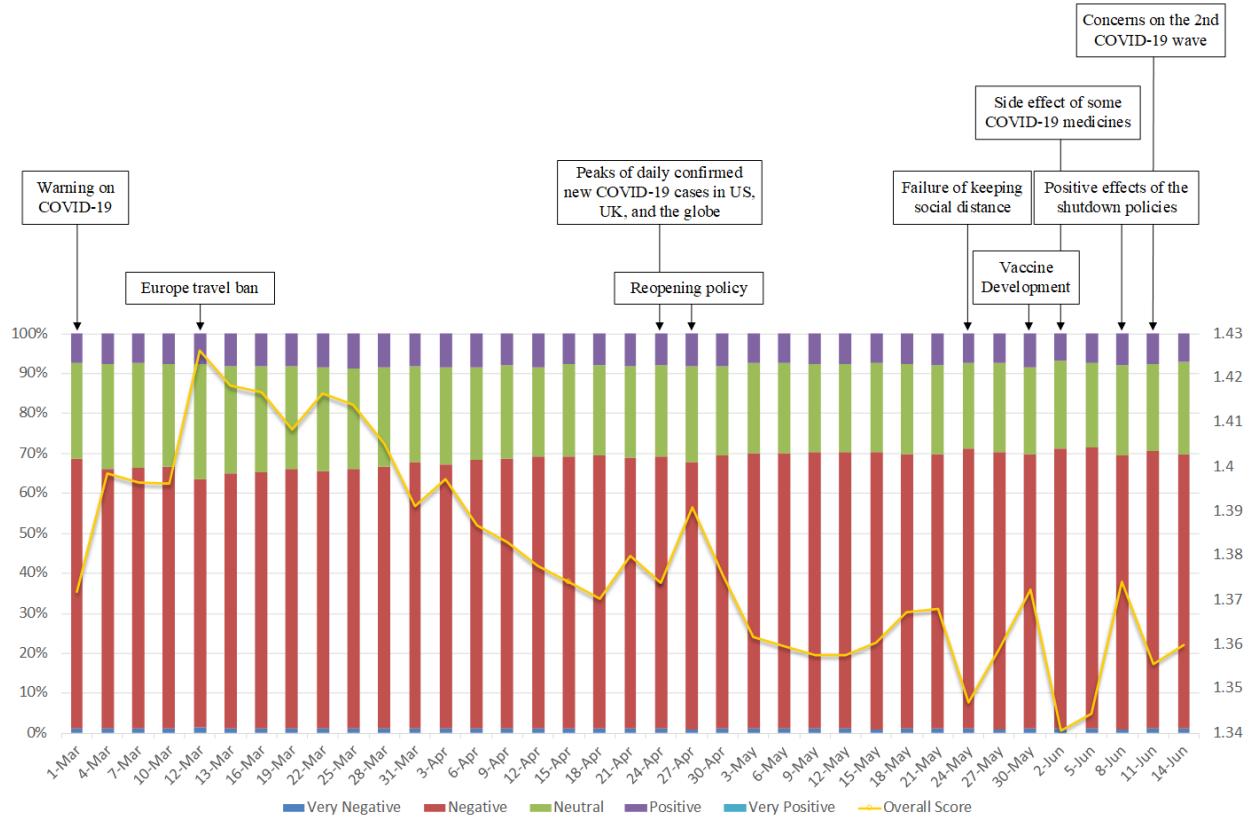


Fig. 5. Lower part: The compositions of sentiments and overall sentimental scores based on model MarAprMay. Upper Part: Policies and events with potential impacts on the sentimental changes of the population.

Figure 5 also shows some deep score valleys. Media's warning on COVID-19 may make the majority of the population nervous and cause the valley on March 1st [43]. Reports about the failure of keeping social distance could be the reason behind the valley on May 24th [44]. The controversy of the side effect of some COVID-19 medicines may contribute to the valley on June 2nd [45]. The concerns on the second COVID-19 wave might be the reason to the valley on June 11th [46].

We also examine other potential influences on the overall sentimental scores. Around April 24th, the numbers of daily confirmed new COVID-19 cases of the United States, the United Kingdom, and the globe reached a peak. This may explain the score valley at that point [47]. However, we do not find a strong correlation between score changes and the CARES Act [48].

## 5. Discussion

This work analyzes tweets related to COVID-19 without considering users' locations, primarily due to the lack of this information in most users' profiles. We may still roughly identify the source of major influence according to special events. For example, the sharp peak on March 12th may suggest that the Twitter users from the United States may be the major influence of the sentimental score changes. It is still hard to conduct local sentiment analysis. For example, it is challenging to

investigate the sentimental changes of the stay-at-home order in a specific state of the United States. Our future works will address these issues by utilizing location information in sentiment analysis. Another area needs more efforts is how to combine the sentimental scores of pre-trained and early-trained models. The current strategy is to provide both scores generated by both models to policy makers. A method to combine these two scores will be helpful.

## **6. Conclusion**

In this paper, we conduct the sentiment analysis of related tweets in the Era of COVID-19. After investigating the sentimental scores of tweets and comparing different models, we propose a new strategy that applies the pre-trained, early-trained, and late-trained models together to build a fast, low-cost, and reliable surveillance system for monitoring the sentimental changes of the population during a pandemic. Furthermore, we analyze the sentimental score changes since March 2020 and connect some visible peaks and valleys to the public policies and events. In our future works, we will focus on enhancing our approach by addressing the location information issue and combining scores of different models.

## **Abbreviations**

EBPH: Evidence-based Public Health Policy; SVM: Support Vector Machine; HPV: Human Papillomavirus; RNTN: Recursive Neural Tensor Network; NLP: Natural Language Processing; RNN: Recursive Neural Network; MSE: Mean Squared Error.

## **Ethics approval and consent to participate**

Not applicable

## **Consent for publication**

Not applicable

## **Availability of data and materials**

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

## **Competing interests**

The authors declare that they have no competing interests.

## **Funding**

Not applicable

## **Authors' contributions**

YW designed the study and drafted the manuscript. MT conducted the data collection and analysis. All authors read and approved the final manuscript.

## Acknowledgements

Not applicable

## Author details

<sup>1</sup>Department of Computer Science and Engineering, University of Tennessee at Chattanooga, Chattanooga, TN 37403, United States.

## References

1. Anderson LM, Brownson RC, Fullilove MT, Teutsch SM, Novick LF, Fielding J, et al. Evidence-based public health policy and practice: Promises and limits. *Am J Prev Med.* 2005;28(5 SUPPL.):226–30.
2. Brownson RC, Fielding JE, Maylahn CM. Evidence-based public health: A fundamental concept for public health practice. *Annu Rev Public Health.* 2009;30:175–201.
3. Brownson RC, Chriqui JF, Stamatakis KA. Understanding evidence-based public health policy. *Am J Public Health.* 2009;99:1576–83.
4. Vanagas G, Bala M, Lhachimi SK. Editorial: Evidence-based public health 2017. *Biomed Res Int.* 2017;2017:1–2.
5. Kohatsu ND, Robinson JG, Torner JC. Evidence-based public health: An evolving concept. *Am J Prev Med.* 2004;27:417–21.
6. Aarons GA, Hurlburt M, Horwitz SMC. Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Adm Policy Ment Heal Ment Heal Serv Res.* 2011;38:4–23.
7. Gibbert WS, Keating SM, Jacobs JA, Dodson E, Baker E, Diem G, et al. Training the workforce in evidence-based public health: An evaluation of impact among US and international practitioners. *Prev Chronic Dis.* 2013;10:1–12.
8. Carter SM, Rychetnik L, Dietetics PgD, Lloyd B, Kerridge IH, Baur L, et al. Evidence, ethics, and values: A framework for health promotion. *Am Intern Med.* 2011;101:465–72.
9. Lim JH, Cho DW. Specular reflection probability in the certainty grid representation. *J Dyn Syst Meas Control Trans ASME.* 1994;116:512–20.
10. Feinstein AR. Clinical epidemiology, I: The populational experiments of nature and of man in human illness. *Am Intern Med.* 1968;69:807–20.
11. Rychetnik L, Hawe P, Waters E, Barratt A, Frommer M. A glossary for evidence based public health. *J Epidemiol Community Health.* 2004;58:538–45.
12. Huston AC. From research to policy and back. *Child Dev.* 2008;79:1–12.
13. Greenlick MR, Goldberg B, Lopes P, Tallon J. Health policy roundtable - view from the state legislature: Translating research into policy. *Health Serv Res.* 2005;40:337–46.
14. Spangler KJ, Caldwell LL. The implications of public policy related to parks, recreation, and public health: A focus on physical activity. *J Phys Act Health.* 2007;4 Suppl 1:S64–71.
15. McGowan A, Schooley M, Narvosa H, Rankin J, Sosin DM. Symposium on public health law surveillance: The nexus of information technology and public health law. *J Law, Med Ethics.*

- 2003;31(4 SUPPL.):41–2.
16. Al-garadi MA, Khan MS, Varathan KD, Mujtaba G, Al-Kabsi AM. Using online social networks to track a pandemic: A systematic review. *J Biomed Inform*. 2016;62:1–11. doi:10.1016/j.jbi.2016.05.005.
  17. Dunn AG, Mandl KD, Coiera E. Social media interventions for precision public health: Promises and risks. *npj Digit Med*. 2018;1:1–4. doi:10.1038/s41746-018-0054-0.
  18. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: A systematic review. *Am J Public Health*. 2017;107:e1–8.
  19. Zhang H, Wheldon C, Tao C, Dunn AG, Guo Y, Huo J, et al. How to improve public health via mining social media platforms: A case study of human papillomaviruses (HPV). In: *Social Web and Health Research*. 2019. p. 207–31.
  20. Merchant RM, Elmer S, Lurie N. Integrating social media into emergency-preparedness efforts. *N Engl J Med*. 2011;365:289–91.
  21. Ratkiewicz J, Meiss M, Conover M, Gonçalves B, Flammini A, Menczer F. Detecting and tracking political abuse in social media. In: *Proceedings of the International AAAI Conference on Weblogs and Social Media*. 2011. p. 297–304.
  22. Lampis V, De Bie T, Cristianini N. Flu detector-tracking epidemics on twitter. In: Balcázar JL, Bonchi F, Gionis A, Sebag M, editors. *Lecture Notes in Computer Science. Machine Learning*. Springer, Berlin, Heidelberg; 2010. p. 599–602.
  23. Top 10 Twitter statistics – updated May 2020. <https://zephoria.com/twitter-statistics-top-ten/>. Accessed 14 Jun 2020.
  24. Twitter usage statistics. <https://www.internetlivestats.com/twitter-statistics/>. Accessed 14 Jun 2020.
  25. Colditz JB, Chu KH, Emery SL, Larkin CR, James AE, Welling J, et al. Toward real-time infoveillance of twitter health messages. *Am J Public Health*. 2018;108:1009–14.
  26. Liu B, Zhang L. A survey of opinion mining and sentiment analysis. In: Aggarwal CC, Zhai C, editors. *Mining Text Data*. Springer, Boston, MA; 2012. p. 415–62.
  27. Vinodhini G, Chandrasekaran R. Sentiment analysis and opinion mining: A survey. *Int J Adv Res Comput Sci Softw Eng*. 2012;2:282–92.
  28. Bhavitha BK, Rodrigues AP, Chiplunkar NN. Comparative study of machine learning techniques in sentimental analysis. In: *Proceedings of the International Conference on Inventive Communication and Computational Technologies (ICICCT)*. 2017. p. 216–21.
  29. Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: A survey. *WIREs Data Min Knowl Discov*. 2018;8:e1253.
  30. Martínez-Cámara E, Martín-Valdivia MT, Ureña-López LA, Montejo-Ráez AR. Sentiment analysis in Twitter. *Nat Lang Eng*. 2014;20:1–28.
  31. Ji X, Chun SA, Wei Z, Geller J. Twitter sentiment classification for measuring public health concerns. *Soc Netw Anal Min*. 2015;5:1–25.
  32. Du J, Xu J, Song H, Liu X, Tao C. Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets. *J Biomed Semantics*. 2017;8:1–7.
  33. Du J, Xu J, Song HY, Tao C. Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. *BMC Med Inform Decis Mak*. 2017;17 Suppl 2.
  34. Cole-Lewis H, Varghese A, Sanders A, Schwarz M, Pugatch J, Augustson E. Assessing electronic cigarette-related tweets for sentiment and content using supervised machine learning. *J Med Internet Res*. 2015;17:1–11.

35. Daniulaityte R, Chen L, Lamy FR, Carlson RG, Thirunarayan K, Sheth A. "When 'Bad' is 'Good'": Identifying personal communication and sentiment in drug-related Tweets. *JMIR Public Heal Surveill*. 2016;2:e162.
36. Tweepy. <https://www.tweepy.org/>. Accessed 26 Jun 2020.
37. Manning C, Surdeanu M, Bauer J, Finkel J, Bethard S, McClosky D. The Stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. 2014. p. 55–60.
38. Socher R, Perelygin A, Wu JY, Ghuang J, Manning CD, Ng AY, et al. Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013. p. 1631–42.
39. Proclamation—Suspension of entry as immigrants and nonimmigrants of certain additional persons who pose a risk of transmitting 2019 novel coronavirus. <https://www.whitehouse.gov/presidential-actions/proclamation-suspension-entry-immigrants-nonimmigrants-certain-additional-persons-pose-risk-transmitting-2019-novel-coronavirus/>. Accessed 26 Jun 2020.
40. COVID-19 pandemic in the United States. [https://en.wikipedia.org/wiki/COVID-19\\_pandemic\\_in\\_the\\_United\\_States](https://en.wikipedia.org/wiki/COVID-19_pandemic_in_the_United_States). Accessed 26 Jun 2020.
41. Graham BBS. Rapid COVID-19 vaccine development. *Science* (80- ). 2020;368:945–6.
42. Hsiang S, Allen D, Annan-Phan S, Bell K, Bolliger I, Chong T, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*. 2020.
43. Fink S, Baker M. Coronavirus may have spread in U.S. for weeks, gene sequencing suggests. <https://www.nytimes.com/2020/03/01/health/coronavirus-washington-spread.html>. Accessed 26 Jun 2020.
44. Shumaker L. Americans spend holiday at beaches and parks as virus death toll nears 100,000. <https://www.reuters.com/article/us-health-coronavirus-usa/americans-spend-holiday-at-beaches-and-parks-as-virus-death-toll-nears-100000-idUSKBN2300L9>. Accessed 26 Jun 2020.
45. Servick K, Enserink M. A mysterious company's coronavirus papers in top medical journals may be unraveling. <https://www.sciencemag.org/news/2020/06/mysterious-company-s-coronavirus-papers-top-medical-journals-may-be-unraveling>. Accessed 26 Jun 2020.
46. Shumaker L, O'Donnell C, Erman M. Fears of second U.S. coronavirus wave rise on worrisome spike in cases, hospitalizations. <https://www.reuters.com/article/us-health-coronavirus-usa-secondwave/fears-of-second-u-s-coronavirus-wave-rise-on-worrisome-spike-in-cases-hospitalizations-idUSKBN23I3DI>. Accessed 26 Jun 2020.
47. JHU CSSE COVID-19 Dataset. <https://github.com/CSSEGISandData/COVID-19>. Accessed 26 Jun 2020.
48. CARES Act. [https://en.wikipedia.org/wiki/CARES\\_Act](https://en.wikipedia.org/wiki/CARES_Act). Accessed 26 Jun 2020.

# Figures

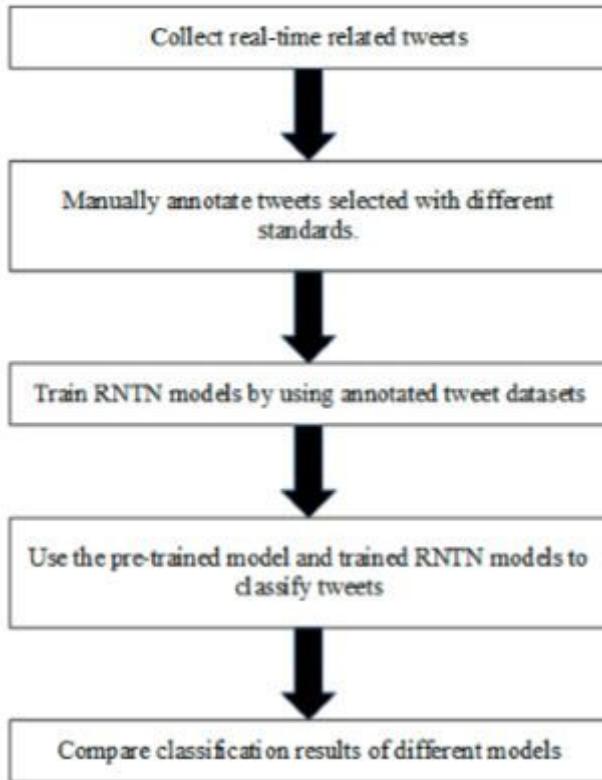


Figure 1

The brief pipeline of investigating related tweets and comparing different models.

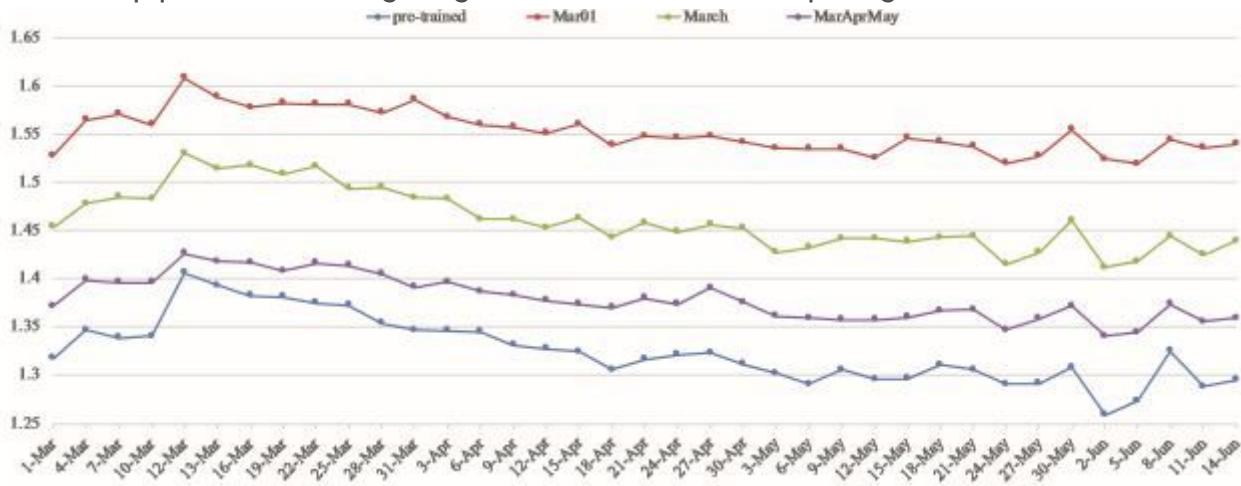
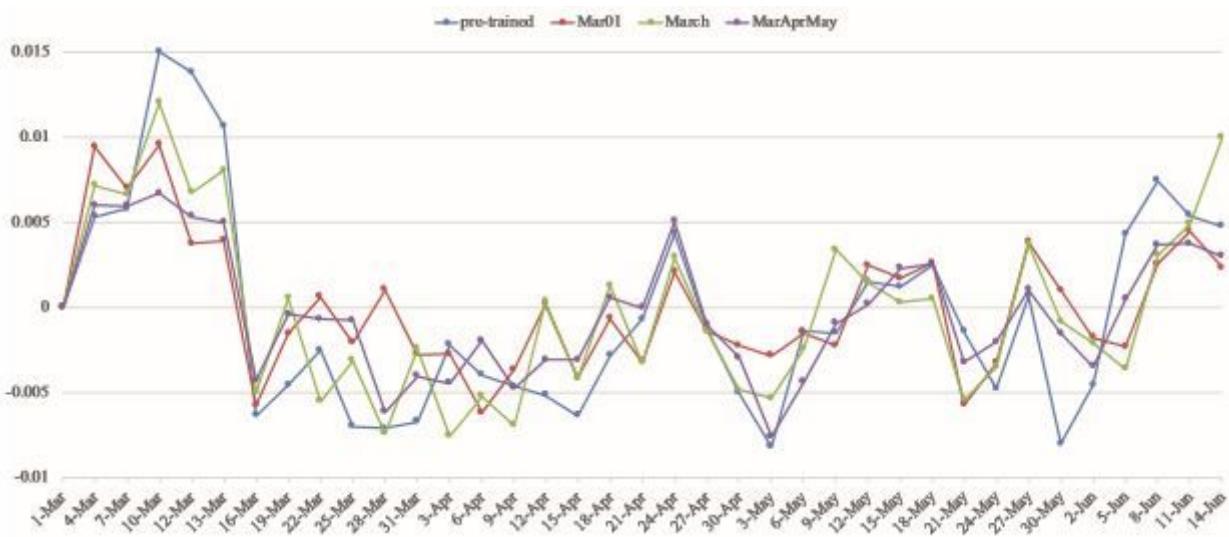


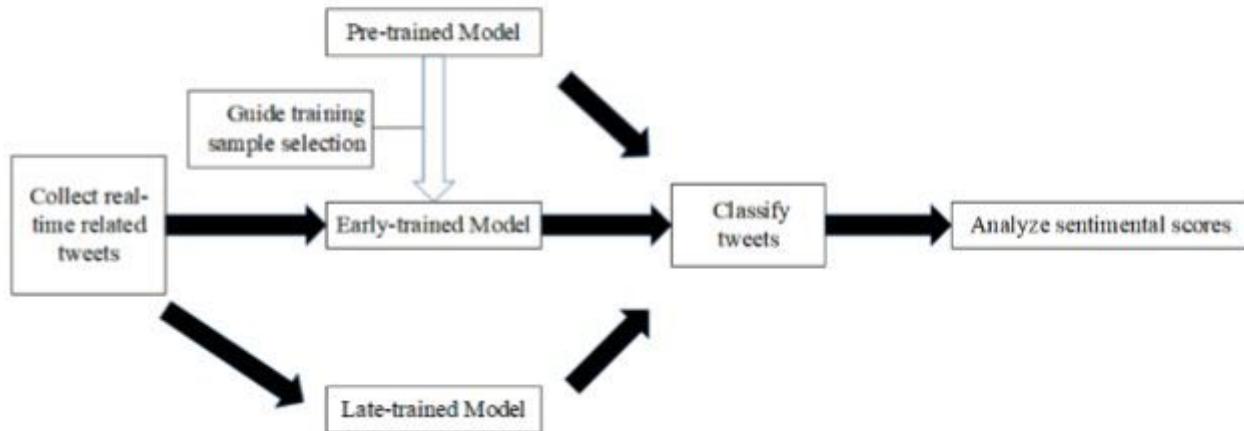
Figure 2

The overall sentimental scores of the pre-trained model, model Mar01, model March, and model MarAprMay.



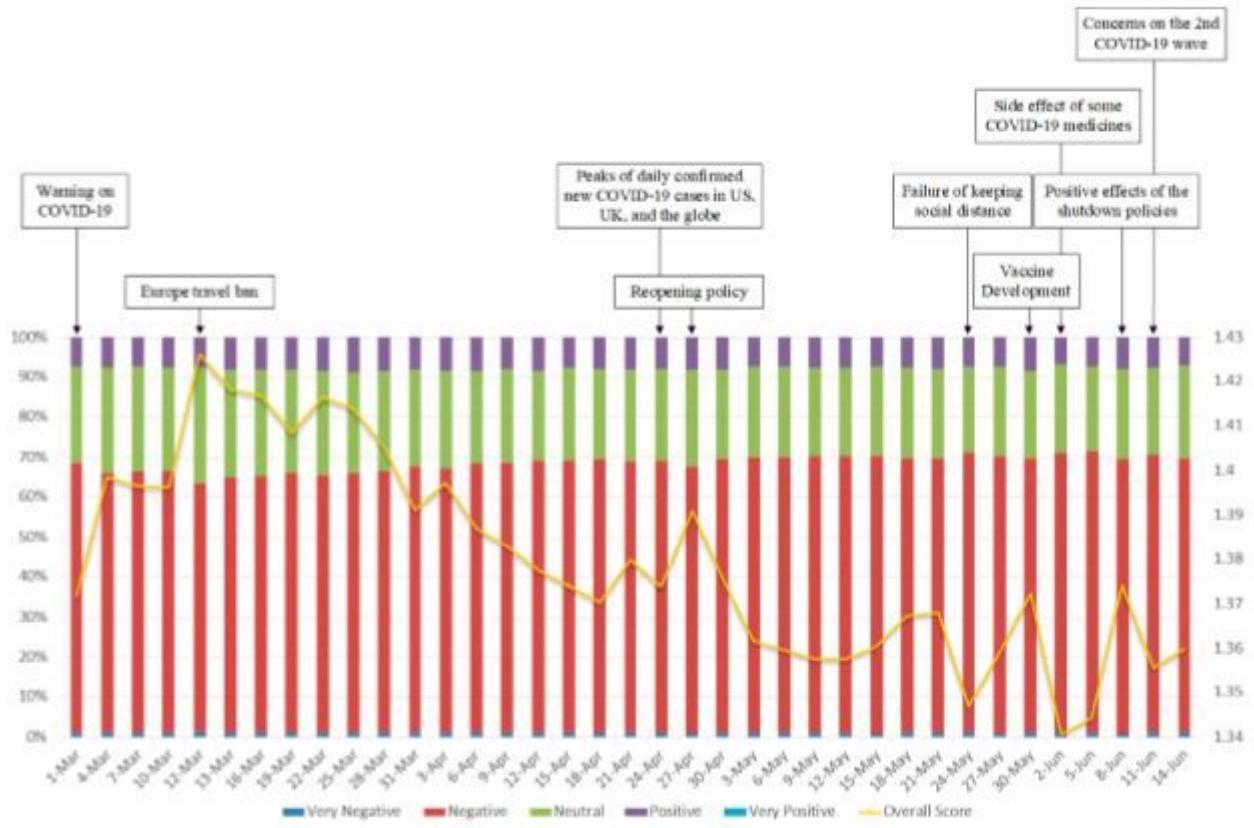
**Figure 3**

The smoothed score changes of the pre-trained model, model Mar01, model March, and model MarAprMay.



**Figure 4**

The pipeline of the proposed strategy for the surveillance system. This system uses the pre-trained and early-trained models in the early stage and the late-trained model in the late stage.



**Figure 5**

Lower part: The compositions of sentiments and overall sentimental scores based on model MarAprMay.  
Upper Part: Policies and events with potential impacts on the sentimental changes of the population.